JOURNAL OF VIROLOGY, Apr. 1990, p. 1487–1497
0022-538X/90/041487-11$02.00/0

Vol. 64, No. 4

Evolution of the Nucleoprotein Gene of Influenza A Virus

OWEN T. GORMAN,* WILLIAM J. BEAN, YOSHIHIRO KAWAOKA, AND ROBERT G. WEBSTER

Department of Virology and Molecular Biology, St. Jude Children's Research Hospital, 332 North Lauderdale, P.O. Box 318, Memphis, Tennessee 38101

Received 13 September 1989/Accepted 19 December 1989

Nucleotide sequences of 24 nucleoprotein (NP) genes isolated from a wide range of hosts, geographic regions, and influenza A virus serotypes and 18 published NP gene sequences were analyzed to determine evolutionary relationships. The phylogeny of NP genes was determined by a maximum-parsimony analysis of nucleotide sequences. Phylogenetic analysis showed that NP genes have evolved into five host-specific lineages, including (i) Equine/Prague/56 (EQPR56), (ii) recent equine strains, (iii) classic swine (H1N1 swine, e.g., A/Swine/Iowa/15/30) and human strains, (iv) gull H13 viruses, and (v) avian strains (including North American, Australian, and Old World subgroups). These NP lineages match the five RNA hybridization groups identified by W. J. Bean (Virology 133:438–442, 1984). Maximum nucleotide differences among the NPs was 18.5%, but maximum amino acid differences reached only 10.8%, reflecting the conservative nature of the NP protein. Evolutionary rates varied among lineages; the human lineage showed the highest rate (2.54 nucleotide changes per year), followed by the Old World avian lineage (2.17 changes per year) and the recent equine lineage (1.22 changes per year). The per-nucleotide rates of human and avian NP gene evolution (1.62×10^{-3} to 1.39×10^{-3} changes per year) are lower than that reported for human NS genes (2.0×10^{-3} changes per year; D. A. Buonagurio, S. Nakada, J. D. Parvin, M. Krystal, P. Palese, and W. M. Fitch, Science 232:980–982, 1986). Of the five NP lineages, the human lineage showed the greatest evolution at the amino acid level; over a period of 50 years, human NPs have accumulated 39 amino acid changes. In contrast, the avian lineage showed remarkable conservatism; over the same period, avian NP proteins changed by 0 to 10 amino acids. The specificity of the H13 NP in gulls and its distinct evolutionary separation from the classic avian lineage suggests that H13 NPs may have a large degree of adaptation to gulls. The presence of avian and human NPs in some swine isolates demonstrates the susceptibility of swine to different virus strains and supports the hypothesis that swine may serve as intermediates for the introduction of avian influenza virus genes into the human virus gene pool. EQPR56 is relatively distantly related to all other NP lineages, which suggests that this NP is rooted closest to the ancestor of all contemporary NPs. On the basis of estimation of evolutionary rates from nucleotide branch distances, current NP lineages are at least 100 years old, and the EQPR56 NP is much older. The ubiquitous nature of hemagglutinin and neuraminidase gene subtypes in the avian gene pool has been interpreted as evidence of an avian origin for influenza A virus. Our results are consistent with this hypothesis. We found that (i) older nonavian NP genes are more avianlike than recent NPs, (ii) recent NP genes in the five lineages are more distant from NPs in other lineages than from the most primitive avian NPs, (iii) nonavian NPs share more primitive avian characters than derived characters, and (iv) Old World avian NP proteins appear to be in evolutionary stasis while nonavian NPs are evolving rapidly away from avianlike ancestral NP proteins. These results suggest independent evolution of NPs from an avian ancestor.

Influenza A viruses infect a variety of avian and mammalian hosts, including humans, pigs, horses, poultry, wild waterfowl, shorebirds, whales, and seals (for a review, see Hinshaw and Webster [11]). Migratory waterfowl and humans have been recognized as major host reservoirs for influenza viruses (11). Recently, Kawaoka et al. (14) presented evidence suggesting that shorebirds and gulls comprise an additional gene pool for influenza viruses.

The results from a number of studies suggest that the nucleoprotein (NP) gene may be involved in the maintenance of host-specific gene pools. Scholtissek et al. (24), Tian et al. (28), and Snyder et al. (26) have implicated the NP gene in determining host range by showing that reassortants with experimentally substituted NP genes have reduced viability. This suggests that NPs are adapted to specific hosts. Scholtissek et al. (25) have shown by RNA hybridization that human NP genes can be readily distinguished from avian and equine NPs. Also using RNA hybridization techniques, Bean (1) identified five host-specific NP groups: equine I (Equine/Prague/56 [EQPR56]), equine II (recent equine), swine plus human, H13 gull, and avian. On the basis of six nucleotide sequences, Buckler-White and Murphy (4) identified human and avian NPs as distinct classes. Gammelin et al. (9) examined 16 NP gene sequences and supported the division of NPs into human and avian classes. Collectively, these studies suggest that NPs can be divided into a minimum of two host-specific classes.

Because the NP gene codes for an internal protein, it is probably not subjected to strong selection pressure by the host immune system. Instead, evolution in the NP gene may reflect host-specific adaptation. The potential host-specific nature of the NP gene makes it a good candidate for the investigation of evolutionary relationships of influenza A viruses. In the study reported here, we present the results of an analysis of 24 influenza A virus NP gene sequences and 18 other NP sequences available from the literature. These NPs represent a wide variety of influenza virus serotypes, hosts, and geographic regions. From our analysis we propose evolutionary relationships among NP genes and address the issue of host specificity.

* Corresponding author.

TABLE 1. Influenza virus strains used in phylogenetic analyses

Strain	Abbreviation	Source or reference
A/Wilson-Smith/33 (H1N1)	WS33	This report
A/Swine/Iowa/15/30 (H1N1)	SWIA30	This report
A/Swine/Tennessee/24/77 (H1N1)	SWTN77	This report
A/Swine/Netherlands/12/85 (H1N1)	SWNED85	This report
A/Equine/London/1416/73 (H7N7)	EQLON73	This report
A/Equine/Kentucky/2/86 (H3N8)	EQKY86	This report
A/Equine/Tennessee/5/86 (H3N8)	EQTN86	This report
A/Gull/Astrakhan/227/84 (H13N6)	GULAST84	This report
A/Gull/Maryland/704/77 (H13N6)	GULMD77	This report
A/Gull/Maryland/1824/78 (H13N9)	GULMD78	This report
A/Gull/Maryland/1815/79 (H13N6)	GULMD79	This report
A/Gull/Minnnesota/1845/80 (H13N6)	GULMN80	This report
A/Gull/Massachusetts/26/80 (H13N6)	GULMA80	This report
A/Duck/New Zealand/31/76 (H4N6)	DKNZ76	This report
A/Grey Teal/Australia/2/79 (H4N4)	GTAUS79	This report
A/Duck/Czeckoslovakia/56 (H4N6)	DKCZ56	This report
A/Duck/Ukraine/2/60 (H11N8)	DKUK60	This report
A/Budgerigar/Hokkaido/1/77 (H4N6)	BDGRHO77	This report
A/Ruddy Turnstone/NJ/47/85 (H4N6)	RTNJ85	This report
A/Whale/Maine/328/84 (H13N2)	WHALEM84	This report
A/Chicken/Pennsylvania/1/83 (H5N2)	CKPENN83	This report
A/Turkey/Minnesota/833/80 (H4N2)	TYMN80	This report
A/Mallard/Astrakhan/244/82 (H?N6)	MLDAST82	This report
A/Tern/South Africa/61 (H5N3)	TERNSA61	This report
A/FPV/Rostock/34/Giessen (H7N1)	FPV34	Mandler and Scholtissek (19)
A/PR/8/34 (H1N1)	PR8-34	Winter and Fields (29)
A/NT/60/68 (H3N2)	NT60-68	Huddleston and Brownlee (12)
A/Hong Kong/5/83 (H3N2)	HK83	Gammelin et al. (9)
A/Udorn/307/72 (H3N2)	UDORN72	Buckler-White and Murphy (4)
A/Mallard/NY/6750/78 (H2N2)	MLRDNY78	Buckler-White and Murphy (4)
A/Chicken/Germany N/49 (H10N7)	CKGER49	Reinhardt and Scholtissek (22)
A/Mink/Sweden/84 (H10N4)	MINKSW84	Reinhardt and Scholtissek (22)
A/Parrot/Ulster/73 (H7N1)	PARU73	Steuler et al. (27)
A/Duck/Hong Kong/7/75 (H3N2)	DKHK75	Gammelin et al. (9)
A/Duck/Bavaria/2/77 (H1N1)	DKBAV77	Gammelin et al. (9)
A/Swine/1976/31 (H1N1)	SW31	Gammelin et al. (9)
A/Swine/Hong Kong/127/82 (H3N2)	SWHK82	Gammelin et al. (9)
A/Swine/Hong Kong/6/76 (H3N2)	SWHK76	Gammelin et al. (9)
A/Swine/Germany/2/81 (H1N1)	SWGER81	Gammelin et al. (9)
A/Equine/Miami/63 (H3N8)	EQMI63	Gammelin et al. (9)
A/Equine/Prague/1/56 (H7N7)	EQPR56	Gammelin et al. (9)
B/Lee/40	B-Lee40	Briedis and Tobin (3)

MATERIALS AND METHODS

Virus strains. A total of 24 viral isolates (Table 1) were selected from the repository at St. Jude Children's Research Hospital, Memphis, Tenn. The isolates were selected to represent a wide spectrum of geographic locations, hosts, and dates of isolation to complement 18 gene sequences from the literature and data bank sources (Table 1).

Molecular cloning of the NP genes. Viruses were grown in 11-day-old embryonated eggs and RNA was extracted as described by Bean et al. (2). Full-length segment 5 genes were cloned by the procedure of Jones et al. (13) and Winter et al. (30) for other influenza A virus genes. To summarize, a 12-mer oligodeoxynucleotide primer (5′ AGCAAAAGC AGG) was phosphorylated with T4 polynucleotide kinase and first-strand cDNA was synthesized from viral RNA template by using avian myeloblastosis virus reverse transcriptase. Second-strand synthesis was done with a phosphorylated 13-mer oligodeoxynucleotide primer (5′ AGTAG AAACAAGC) and *Escherichia coli* DNA polymerase I (Klenow fragment). Full-length double-stranded cDNA was blunt-end ligated into the *Pvu*II site of vector pATX (a derivative of pAT 153, courtesy of Clayton Naeve, Molecular Resource Center, St. Jude Children's Research Hospital).

Nucleotide sequence determination. The nucleotide sequences of NP cDNAs were determined by using the dideoxynucleotide-chain termination method (23). Oligodeoxynucleotide primers (synthesized by the Molecular Resource Center) were annealed to double-stranded template DNA denatured with NaOH as described by Chen and Seeburg (6) and extended with modified T7 DNA polymerase (Sequenase; U.S. Biochemical, Cleveland, Ohio). The reaction products were separated on 6% polyacrylamide—7 M urea 0.4-mm gels. Ambiguities were resolved by sequencing the complementary strand or the original viral RNA with avian myeloblastosis virus reverse transcriptase. For all virus strains, at least one-third of the NP gene was sequenced on both strands. Two or more clones were available and sequenced for 15 of the 24 virus isolates. The sequences of oligonucleotides used as primers are available upon request.

Sequence analysis. The Intelligenetics (Palo Alto, Calif.) software package was used for analysis and translation of nucleotide sequence data. The B/Lee/40 NP nucleotide

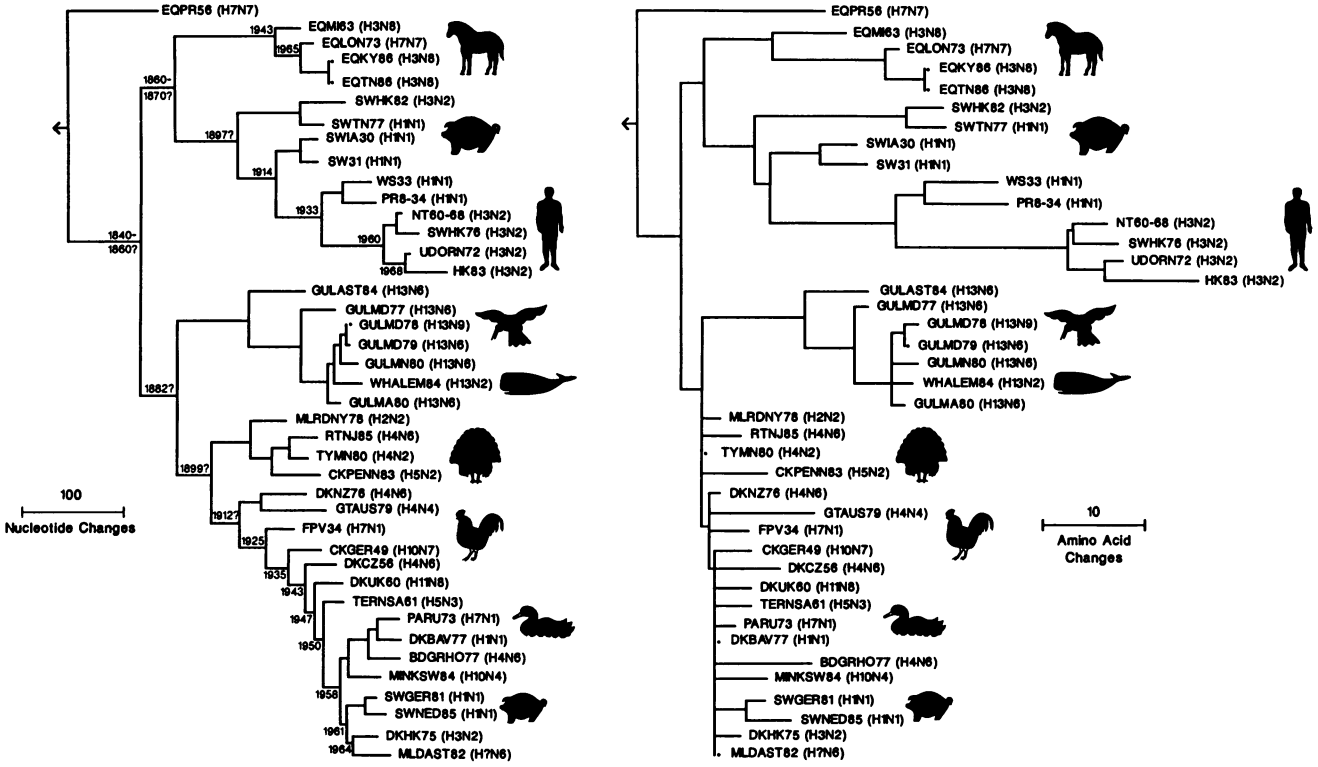

FIG. 1. Phylogenetic trees for influenza A virus NP genes. (Left) Nucleotide tree rooted to B/Lee/40 NP. Sequences of 42 NP genes (24 from this study, 18 from published sources) were analyzed with PAUP (D. Swofford, Illinois Natural History Survey), which uses a maximum-parsimony algorithm. The tree represents the first of six equal-length trees generated by PAUP with MULPARS, SWAP= GLOBAL, and HOLD=10 options. The arrow indicates the direction of the B/Lee/40 NP from the root node. With the exclusion of the B/Lee/40 outgroup, the number of variable characters represented is 641, the total tree length is 2,289 nucleotide changes, and the consistency index (proportion of changes due to forward mutations) is 0.389. Horizontal distance is proportional to the minimum number of nucleotide differences to join nodes and NP gene sequences. Vertical lines are for spacing branches and labels. Dates for hypothetical ancestor nodes were derived by dividing branch distance by evolutionary rate estimates (see Fig. 5). Question marks denote date estimates beyond the immediate ancestor node of the earliest isolates. (Right) Amino acid tree. The tree was generated by using the topology option of PAUP to make predicted amino acid sequences conform to the topology of the nucleotide tree (left). The number of variable characters represented is 141, the total tree length is 275 amino acid changes, and consistency is 0.680. Strain abbreviations used in these trees are listed in Table 1.

sequence (1,841 bases [3]) was aligned with the influenza A virus NPs by use of the Needleman-Wunsch pairwise alignment algorithm. Alignment was accomplished by eight deletions (276 bases) in the first half of the gene. Phylogenetic analysis of sequence data was performed with the PAUP software package version 2.4.1 (David Swofford, Illinois Natural History Survey, Champaign, Ill.). PAUP employs the maximum-parsimony method to generate phylogenetic trees. The shortest (most parsimonious) trees were found by implementing the MULPARS, SWAP=GLOBAL, and HOLD=10 options of PAUP. The PRINTD option of PAUP provided difference matrices for sequence data. Tree length is measured in steps which are equivalent to nucleotide changes for nucleotide trees and amino acid changes for amino acid trees. The total tree length is the sum of all branch lengths.

RESULTS

Analysis of nucleotide sequences. Each of the 24 cloned NP genes comprised 1,565 nucleotides and a single open reading frame which spanned positions 46 through 1539 and coded for a polypeptide of 498 amino acids. No insertions or deletions were observed in any of the sequences. Nucleotide sequences of the 24 cloned NP genes are not presented here but are available from GenBank (accession numbers M30746 to M30769).

A phylogenetic analysis of 41 influenza A virus NP nucleotide sequences (from the 24 cloned genes and 17 published sequences) is presented as an evolutionary tree rooted to an aligned B/Lee/40 sequence (Fig. 1, left). With the exclusion of the B/Lee/40 sequence, PAUP found six equal-length trees which were 2,289 steps (nucleotide changes) long. The next-shortest trees were 11 steps longer. The six equal-length trees varied in combinations of two possible placements of GULMD79 and MINKSW84 with adjacent terminal branches. We chose that tree (tree 1) with a branching order for GULMD79 and MINKSW84 that was most congruent with the sequence of isolation dates (Fig. 1, left).

When conducting a phylogenetic analysis, it is desirable to use an outgroup (a closely related taxon that shares a common ancestor) to serve as a reference or root for the group of interest (in our case, influenza A virus NPs). If the outgroup is closely related it can more accurately resolve differences between lineages within the group of interest. If the outgroup is too distantly related, its value as a reference is diminished. Initially, we included two aligned influenza B

virus NP sequences (B/Lee/40 and B/Singapore/79 [18]) and an aligned influenza C virus NP sequence (C/Calif/78 [20]) as outgroup sequences in our analysis. Of the three outgroup NPs, the NP of B/Lee/40 was the closest to influenza A virus NPs and was chosen as the outgroup. The influenza C virus NP was eliminated from consideration because it was much more distant from influenza A virus NPs. Structural similarities found between influenza A and B virus hemagglutinins (17) suggest a close evolutionary relationship between the two viruses and support the choice of influenza B virus NP as the outgroup. The NP of B/Lee/40 is separated from the root of influenza A virus NPs by a minimum of 965 nucleotide changes (this does not include the eight deletions totaling 276 nucleotides made in the B/Lee/40 NP sequence to align it with influenza A virus NPs). At 1,053 steps, the NP of EQPR56 is the closest to that of B/Lee/40, and the next closest NP, that of MLRDNY78, is 1,191 steps away.

The phylogenetic tree shows that NP genes have evolved into five divergent, host-specific lineages rooted at the deep forks of the tree (Fig. 1, left). The first lineage is that of EQPR56 NP, which forms a sister group to all other influenza A virus NPs. The relative closeness to the B/Lee/40 root and the distance to other NPs indicates that EQPR56 has the most primitive NP; the closest NPs in other lineages are 314 to 343 steps away, and contemporary avian and human NPs are more than 400 steps away. The next fork in the tree represents a separation of H13 gull and avian lineages from recent equine and swine plus human lineages; the closest NPs between the two sides of the tree (EQMI63 and MLRDNY78) are separated by 311 steps. The closest NPs between the equine and swine plus human lineages (EQMI63 and SW31) are separated by 270 steps. The NPs of the H13 gull lineage are distinct from the other avian NPs; the closest of these (GULAST84 and MLRDNY78) is separated by 244 steps. With the exception of one NP derived from a whale isolate, all of our H13 NPs were derived from gull isolates.

The recent equine NP lineage contains examples of viruses with H3N8 and H7N serotypes. Although the recent equine EQLON73 strain shares the same serotype (H7N7) as the EQPR56 strain, their NPs are quite distinct. Bean (1) has suggested that recent equine viruses of the H7N7 subtype may be reassortants that have replaced the EQPR56 NP with the recent equine NP. The failure to isolate viruses containing EQPR56-like NPs since 1956 suggests that this lineage may be extinct.

The union of classic H1N1 swine and human NP lineages suggests that the human NPs are derived from a swine ancestor or at least that the two lineages share a common ancestor. In the classic swine lineage, SWHK82 is a reassortant containing avian or early human H3N2 surface protein genes (e.g., Aichi/68 [16]) and a classic H1N1 swine NP. In the human lineage the swine isolate SWHK76 is a typical H3N2 human virus, but the tree suggests that its NP was derived from human NPs of the 1960s.

The avian lineage is a heterogeneous group; NPs were isolated from wild and domestic ducks, shorebirds, turkeys, chickens, parrots, budgerigars, mink, and swine. To simplify discussion we will hereafter refer to this lineage as avian and to the others (including H13 gull) as nonavian. Within the avian lineage, readily identifiable subgroups are defined by geographic region: the North American lineage, which is closest to the ancestral avian NP (e.g., MLRDNY78); the Australia-New Zealand lineage (e.g., DKNZ76); and the Old World lineage, with FPV34 as its most primitive member.

Over all NPs, the maximum absolute number of nucleotide differences were found between the recent (most derived) NPs of the swine plus human, recent equine, and H13 gull lineages (257 to 289 differences or 16.4 to 18.5%; Fig. 2). The oldest NPs of these lineages were less divergent (231 to 252 differences or 14.8 to 16.1%) and were more similar to older, more primitive avian NPs, e.g., FPV34 (197 to 222 differences or 12.6 to 14.2%).

In contrast to absolute differences, branch distances shown in the phylogenetic tree (Fig. 1, left) account for reverse mutations and thus provide more accurate estimates of evolutionary distance. It is readily apparent that the most recent NPs at the terminal branches of nonavian lineages are the most divergent NPs (they are separated from each other by a distance of 416 to 535 steps) but all are closer to the root of the avian lineage (263 to 382 steps; root dated 1899 in Fig. 1, left). Similarly, the oldest NPs of the nonavian lineages are closer to the avian root (201 to 249 steps) than to each other (270 to 376 steps). Overall, nonavian NPs share more primitive (avian) characters than derived (nonavian) characters. This pattern of relationship can be explained by divergent evolution of nonavian NP genes from a common avian origin. The same type of pattern appears within the avian lineage; e.g., the most divergent avian NPs, CKPENN83 and BDGRHO77, are separated by a branch distance of 311 steps but they are only 111 and 200 steps, respectively, from the avian root. Relationships among avian NP branches suggest evolution from a common ancestor.

A low consistency index of 0.389 and the much higher number of nucleotide changes predicted by the phylogenetic tree in comparison with the absolute differences (Fig. 1 versus Fig. 2) indicates that reverse mutations are common in the evolution of the NP gene (the consistency index is a measure of the proportion of nucleotide changes due to forward mutations in phylogenetic trees). These reversals are probably related to constraints on NP amino acids; that is, silent mutations at the second and third positions of nucleotide codons are saturated in some lineages and reversals are selected for in order to maintain ancestral amino acid sequences.

Analysis of amino acid sequences. To evaluate the product of nucleotide evolution and the host-specific nature of NP genes, we translated the open reading frames of the 41 influenza A virus nucleotide sequences. The absolute differences in deduced amino acid sequences were much less than expected, given the large number of nucleotide differences, which indicates that the NP protein is conserved, especially within avian NPs (Fig. 2). Within mammalian lineages, NPs show four to eight nucleotide changes for each amino acid change between isolates, but in the H13 lineage the range is 4 to 15 changes and in the avian lineage the range is 6 to >40 changes. The most divergent amino acid sequences are the human HK83 and H13 GULMN80 NPs (54 differences or 10.8%). The pattern of differences is the same as the nucleotide data; the greatest differences occur among mammalian and H13 gull lineages, and the oldest NPs are more similar to avian NPs.

Since amino acid evolution is determined by nucleotide evolution, we used the topology option of PAUP to generate an amino acid sequence tree that conforms to the branching pattern of the nucleotide tree (Fig. 1, right). This approach allows direct comparison of the nucleotide and amino acid trees for differences in the effect of genetic changes among the lineages. The tree shows that at the protein level the various NP lineages are divergent and evolving at different rates. The avian lineage is the most primitive; several avian NPs (TYMN80, DKBV77, MLDAST82) are separated from

DIFFERENCES IN NUCLEOTIDE SEQUENCES

	VIRUS	EP	EQ				SW				HUM						H13							NA				ANZ		OW												
	STRAIN	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41
1	EQPR56	0	262	267	272	272	258	264	232	231	260	251	262	267	270	275	256	265	267	269	261	272	277	238	238	234	238	222	225	218	219	226	223	213	219	234	214	221	208	220	218	212
2	EQM163	33	0	64	80	80	269	260	247	247	249	252	263	260	270	281	243	246	253	250	256	249	251	201	202	196	213	204	218	200	201	207	196	199	203	211	204	206	199	207	208	194
3	EQLON73	35	19	0	36	36	264	260	243	240	246	245	259	257	264	277	255	246	256	255	259	252	258	223	220	217	230	223	235	232	223	226	215	218	226	233	226	226	218	227	222	211
4	EQKY86	40	22	5	0	0	268	264	250	249	250	250	263	260	269	278	257	250	256	255	259	256	260	228	224	225	234	226	236	237	232	233	219	224	234	242	240	234	228	235	229	217
5	EQTN86	40	22	5	0	0	268	264	250	249	250	250	263	260	269	278	257	250	256	255	259	256	260	228	224	225	234	226	236	237	232	233	219	224	234	242	240	234	228	235	229	217
6	SWHK82	37	34	36	40	40	0	67	181	182	207	209	227	233	235	248	273	272	267	267	269	270	270	251	239	244	255	230	254	243	233	243	240	232	234	234	229	245	233	235	237	248
7	SWTN77	40	35	39	43	43	8	0	153	150	180	185	212	222	216	235	269	274	275	275	276	273	279	250	244	249	256	237	246	249	233	248	246	234	249	248	244	252	242	242	245	248
8	SWIA30	32	25	29	33	33	24	25	0	27	113	125	164	174	172	196	230	231	241	241	242	236	245	210	202	202	224	197	202	206	202	203	208	194	210	214	209	213	200	210	202	204
9	SW31	33	25	30	33	33	23	26	7	0	114	125	154	166	164	192	227	232	242	242	245	237	248	206	199	201	219	202	204	206	199	201	209	196	213	216	210	213	200	210	205	204
10	WS33	42	37	40	44	44	38	38	25	26	0	60	127	142	136	165	252	252	257	257	260	253	265	217	213	212	236	214	227	222	213	227	224	210	221	229	221	222	215	224	215	219
11	PR834	41	41	44	46	46	38	39	28	29	15	0	127	144	139	166	262	251	262	262	265	260	274	225	227	221	245	216	229	220	213	225	223	208	221	232	222	221	211	220	207	217
12	NT60-68	48	45	45	47	47	45	44	41	42	29	30	0	30	49	78	264	264	273	272	275	271	283	245	242	246	248	237	253	249	244	246	240	234	240	246	244	235	229	236	241	240
13	SWHK76	44	41	43	45	45	42	43	39	40	29	30	7	0	63	92	264	266	274	274	277	273	283	241	245	247	245	235	251	247	246	243	238	240	240	249	246	234	234	240	243	241
14	UDORN72	48	45	45	47	47	46	47	41	42	32	32	10	11	0	49	265	264	270	270	273	266	278	254	253	253	267	246	257	253	249	257	251	251	253	260	255	250	245	251	255	252
15	HK83	50	49	49	51	51	50	49	45	46	38	38	15	18	11	0	271	277	282	282	284	278	289	270	266	268	276	255	269	266	259	269	263	257	258	260	265	264	257	263	267	259
16	GULAST84	37	30	32	37	37	39	41	27	28	41	43	50	46	48	52	0	143	141	140	151	139	154	215	216	203	209	196	227	199	210	215	207	196	211	219	211	213	202	217	214	213
17	GULMD77	39	28	30	35	35	38	41	26	27	39	43	48	44	48	52	13	0	71	69	82	73	90	224	207	201	220	209	218	197	196	209	203	195	213	214	214	213	199	199	200	211
18	GULMD78	40	29	35	40	40	38	41	27	28	40	44	48	43	48	52	16	7	0	4	20	23	35	225	223	211	230	217	232	200	202	211	204	200	220	229	220	214	202	209	206	216
19	GULMD79	39	28	34	39	39	37	40	26	27	39	43	47	42	47	51	16	6	1	0	20	19	35	223	219	209	230	213	232	196	198	209	202	196	216	225	216	210	198	205	202	214
20	GULMN80	41	30	36	41	41	39	42	28	29	41	45	50	45	50	54	18	8	5	4	0	31	49	229	224	217	233	215	230	202	203	214	203	203	219	232	223	215	201	210	205	217
21	GULMA80	39	28	34	39	39	37	40	26	27	39	43	48	43	48	52	16	6	3	2	2	0	39	224	219	210	231	217	233	198	200	209	205	198	218	225	216	218	202	210	203	217
22	WHALEM84	38	27	33	38	38	36	39	25	26	38	42	47	42	47	51	17	7	4	3	5	3	0	229	223	211	234	225	236	209	210	217	208	206	220	229	220	223	206	214	209	218
23	MLRDNY78	27	16	23	28	28	27	28	13	14	27	31	39	34	39	43	19	17	17	16	18	16	15	0	92	82	106	154	167	147	140	146	150	128	152	160	148	146	145	154	146	138
24	RTNJ85	27	18	23	28	28	26	27	14	15	27	31	38	34	38	42	21	19	20	19	21	19	18	6	0	49	91	155	162	142	133	149	147	132	143	153	136	154	145	158	143	144
25	TYMN80	25	14	21	26	26	25	26	11	12	25	29	37	33	37	41	17	15	16	15	17	15	14	2	4	0	82	153	157	127	125	138	137	121	140	150	135	140	137	150	131	137
26	CKPENN83	30	19	24	29	29	32	31	18	19	32	36	40	38	40	40	21	21	22	21	23	21	20	9	11	7	0	178	179	156	154	155	157	146	173	176	170	168	163	172	163	163
27	DKNZ76	23	16	21	26	26	23	26	11	12	25	29	37	33	37	41	18	16	17	16	18	16	15	4	6	2	9	0	118	113	128	135	135	132	146	153	147	148	143	151	138	140
28	GTAUS79	29	22	26	31	31	29	30	18	19	29	33	40	36	40	44	28	26	27	26	28	26	25	13	14	11	18	11	0	138	149	158	147	152	154	161	155	163	152	158	156	150
29	FPV34	28	17	26	31	31	28	29	12	13	26	30	40	36	38	42	20	20	21	20	22	20	19	7	9	5	10	5	14	0	86	95	91	87	103	121	99	115	108	119	108	113
30	CKGER49	27	19	24	29	29	26	25	12	14	26	30	38	34	38	42	21	19	20	19	21	19	18	7	7	5	12	5	13	8	0	84	79	76	96	112	95	90	93	102	95	108
31	DKCZ56	26	19	26	31	31	28	29	15	16	28	32	41	37	41	45	23	21	22	21	23	21	20	8	10	6	13	6	14	9	9	0	65	56	92	110	97	98	94	111	89	90
32	DKUK60	27	17	22	27	27	26	27	13	14	26	30	38	34	38	42	22	18	21	20	22	20	19	7	7	5	12	5	13	6	6	9	0	56	81	101	85	87	81	96	76	91
33	TERNSA61	26	19	22	27	27	26	25	13	14	26	30	37	35	39	41	22	20	21	20	22	20	19	7	7	5	10	5	13	8	6	7	6	0	67	84	69	82	69	86	60	79
34	PARU73	26	18	23	28	28	25	26	12	13	25	29	36	32	36	40	21	19	20	19	21	19	18	6	6	4	11	4	12	7	5	8	5	5	0	61	39	76	58	86	62	80
35	BDGRH077	30	22	30	35	35	30	29	18	19	32	36	42	40	44	46	25	24	24	23	25	23	22	11	13	9	14	9	16	12	12	13	12	10	11	0	68	88	91	114	92	107
36	DKBAV77	24	16	21	26	26	23	24	10	11	23	27	36	32	36	40	19	17	18	17	19	17	16	4	4	2	9	2	10	5	3	6	3	3	2	9	0	80	68	95	72	87
37	MINKSW84	29	19	24	29	29	26	27	15	16	28	32	40	36	41	45	23	21	22	21	23	21	20	9	9	7	12	7	15	10	8	11	8	8	7	14	5	0	78	95	70	85
38	SWGER81	26	19	25	30	30	27	28	14	15	26	30	39	35	39	43	21	19	20	19	21	19	20	8	8	6	13	6	14	9	7	10	7	7	4	13	4	9	0	33	58	70
39	SWNED85	27	20	26	31	31	28	29	17	18	27	31	36	32	36	40	24	22	23	22	24	22	23	11	11	9	16	9	15	12	10	13	10	10	7	16	7	12	5	0	74	83
40	DKHK75	26	18	23	28	28	25	26	12	13	25	29	38	34	38	42	21	19	20	19	21	19	18	6	6	4	11	4	12	7	5	8	5	5	4	11	2	7	6	9	0	61
41	MLDAST82	24	16	21	26	26	23	24	10	11	23	27	36	32	36	40	19	17	18	17	19	17	16	4	4	2	9	2	10	5	3	6	3	3	2	9	0	5	4	7	2	0

DIFFERENCES IN AMINO ACID SEQUENCES

FIG. 2. Difference matrix of nucleotide and amino acid sequences for 41 influenza A virus NP genes. Group divisions correspond to lineages in Fig. 1 as follows: EP, EQPR56; EQ, recent equine; SW, classic H1N1 swine; HUM, human; H13, H13 gull; NA, North American avian; ANZ, Australia and New Zealand avian; OW, Old World avian. Strain abbreviations and sources are listed in Table 1.

the root of influenza A virus NPs by fewer than eight amino acid differences, and other avian NPs are nearly as close. Evolution in the avian lineage is virtually static; for example, although isolated 48 years apart, FPV34 differs by only four amino acid changes from the recent MLDAST82 isolate. Other lineages show significant evolution at the protein level, especially human NPs. Nonavian NPs at terminal branches differ from each other by 47 to 75 amino acid changes, but each is a minimum of 18 steps closer to the avian root (22 to 53 steps to the node above MLRDNY78 in Fig. 1, right). Similarly, the oldest nonavian NPs differ by 29 to 41 amino acid changes but are at least 10 steps closer to the avian root (18 to 25 steps). As with the nucleotide tree, this pattern of evolution away from avianlike NPs suggests an avian origin for all NPs.

A comparison of amino acid sequences among the five major NP lineages shows suites of amino acids unique to each lineage (Fig. 3). To detect patterns of unique (derived) amino acids among the five lineages, we used the most primitive NP sequence, TYMN80 (the NP closest to the root of the tree), as a baseline. This choice is critical; if we used a derived human NP for a baseline, only two groups would be evident, human and nonhuman NPs. The nonhuman NPs would appear to form a group because they have retained primitive avian amino acids that were lost independently in the human NPs. In this situation and nonhuman NPs would be artificially united by primitive rather than derived character states.

Despite more than 300 nucleotide changes among avian NP genes, only two amino acids show a pattern correlated with a group: North American NPs possess unique amino acids at sites 105 and 450. Mandler and Scholtissek (19) have reported a temperature-sensitive mutant of FPV34 in which an amino acid change occurs at position 332. This mutation is located within a 24-residue conserved region for the 41 NPs we analyzed (Fig. 3 and 4). These findings emphasize the conserved nature of the NP protein among avian strains. The swine and mink isolates within the avian group and the whale isolate within the H13 gull group show no unusual patterns of amino acid substitutions, nor convergence with NPs of the mammalian lineages.

Amino acid changes that are correlated with groups are summarized in Fig. 4. In comparison with the avian NPs, the human NPs possess the most fixed amino acid differences (39 in current human NPs) and H13 gull NPs possess the fewest (15 in current H13 NPs). EQPR56, recent equine, and classic swine NPs have 22, 26, and 19 fixed amino acid differences, respectively. Most of these differences are unique to each lineage, indicating that evolution of NPs in each lineage is divergent and independent. For example, even though human and classic swine appear to share a common ancestor on the basis of the nucleotide phylogeny (Fig. 1, left), the oldest isolates share only two unique amino acids. Shared derived residues acquired at a later date must be due to convergence or parallelism. A comparison of accumulated amino acid differences over the past 50 years indicates that evolution at the protein level is about twice as fast in the human lineage (39 versus 19 residues).

Estimation of evoluationary rates. Within the mammalian and avian lineages, the oldest virus isolates are more than 50 years old. The long period over which isolates were obtained provided an opportunity to estimate long-term evolutionary rates for NP genes in different lineages. The evolutionary rate was estimated by plotting the year of isolation for a virus

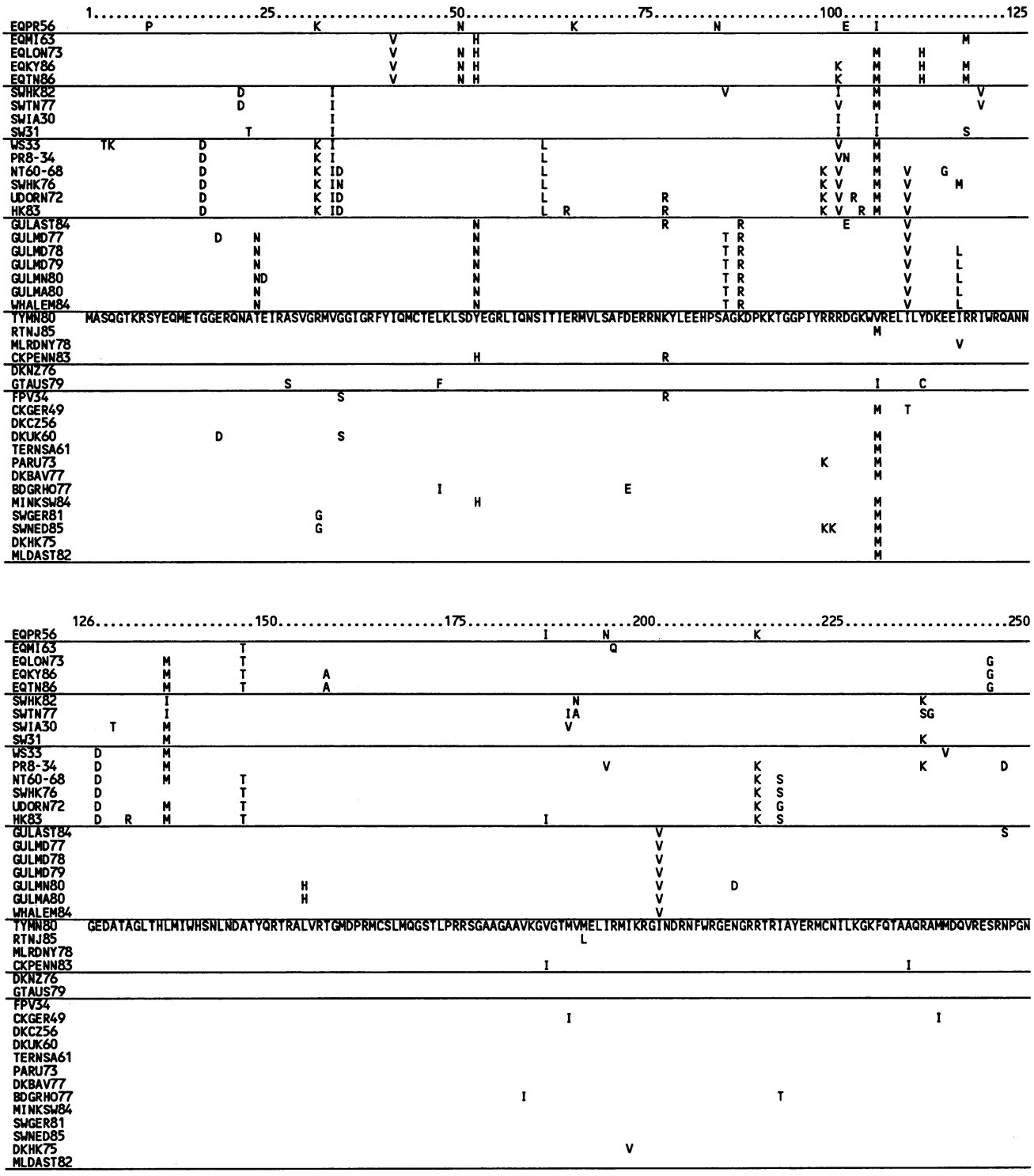

FIG. 3. Predicted amino acid sequences of the 41 influenza A virus NP genes. Amino acid sequence of TYMN80 is written in full, and for other sequences only differences from this baseline are shown. NP groups recognized as lineages in Fig. 1 are separated by lines in the following order (top to bottom): EQPR56, recent equine, classic H1N1 swine, human, H13 gull, North American avian, Australia and New Zealand avian, and Old World avian. Strains represented are listed in Table 1.

against the branch distance to the ancestor node of the lineage; the slope of the regression line for the plotted points equals the number of nucleotide changes per year (Fig. 5). In the human lineage, WS33 and PR8-34 are closest to the ancestor node (Fig. 1) but are outliers (Fig. 5); they appear to have too many derived characters for their age. This situation may be the result of continuous passaging of the viruses before the advent of deep freezers. Regression of the remaining datum points yielded a slope of 2.54 nucleotide changes per year (1.62×10^{-3} changes per nucleotide per year). The

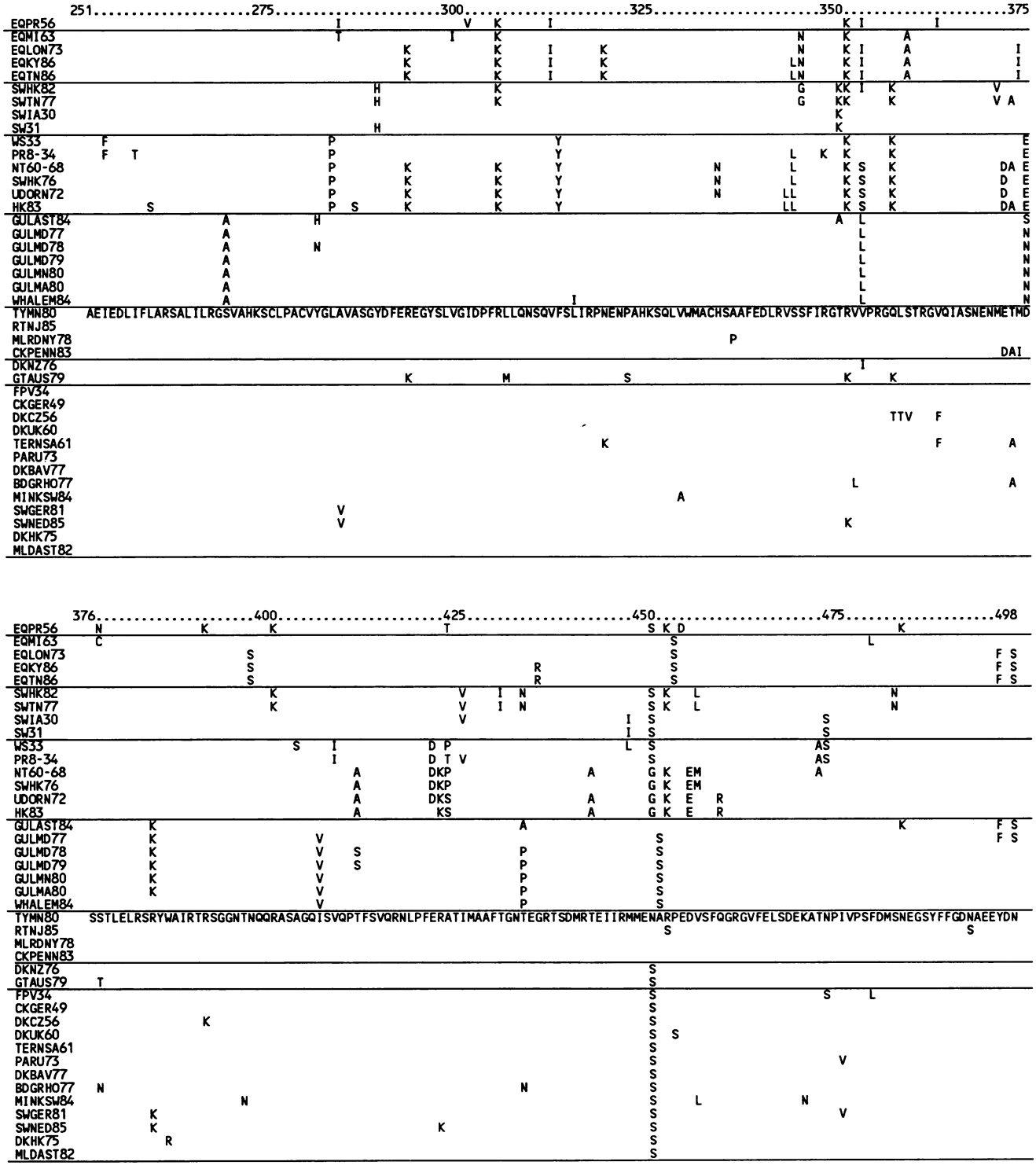

FIG. 3—*Continued.*

estimated evolutionary rate for the Old World avian lineage was slightly lower (2.17 changes per year; 1.39×10^{-3} changes per nucleotide per year) and lower still for the recent equine lineage (1.22 changes per year; 0.78×10^{-3} changes per nucleotide per year). Our values are smaller than the estimates reported for equine hemagglutinin (HA) genes (2.8×10^{-3} substitutions per site per year [7]) and human nonstructural (NS) genes (2×10^{-3} substitutions per site per year [5]). The accuracy of these estimates is improved by a longer period over which isolates are obtained, the inclusion of more isolates, and an absence of large time gaps between isolates. Based on these criteria, our estimate of evolutionary rate is most accurate for the Old World avian lineage. Because of the short time span over which North

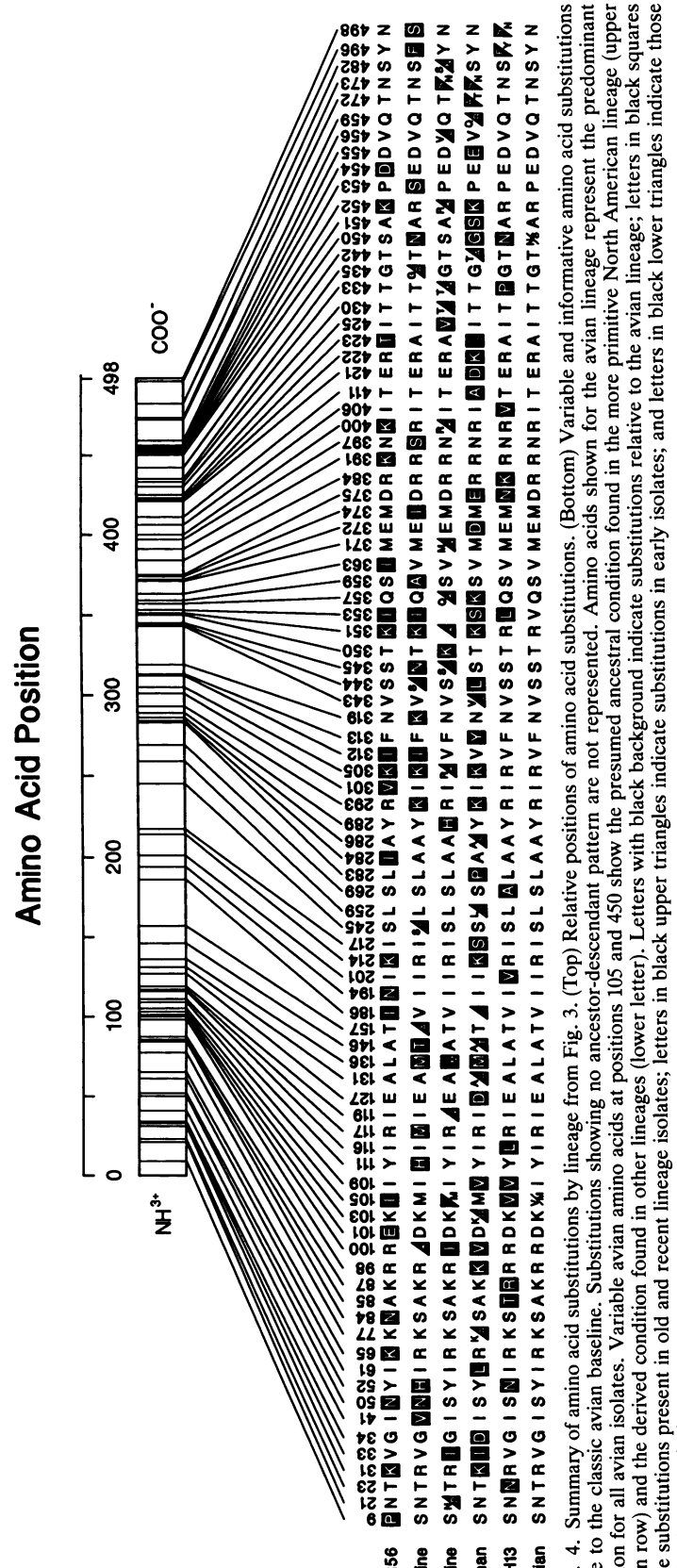

FIG. 4. Summary of amino acid substitutions by lineage from Fig. 3. (Top) Relative positions of amino acid substitutions. (Bottom) Variable and informative amino acid substitutions relative to the classic avian baseline. Substitutions showing no ancestor-descendant pattern are not represented. Amino acids shown for the avian lineage represent the predominant condition for all avian isolates. Variable avian amino acids at positions 105 and 450 show the presumed ancestral condition found in the more primitive North American lineage (upper letter in row) and the derived condition found in other lineages (lower letter). Letters with black background indicate substitutions relative to the avian lineage; letters in black squares indicate substitutions present in old and recent lineage isolates; letters in black upper triangles indicate substitutions in early isolates; and letters in black lower triangles indicate those occurring in recent isolates.

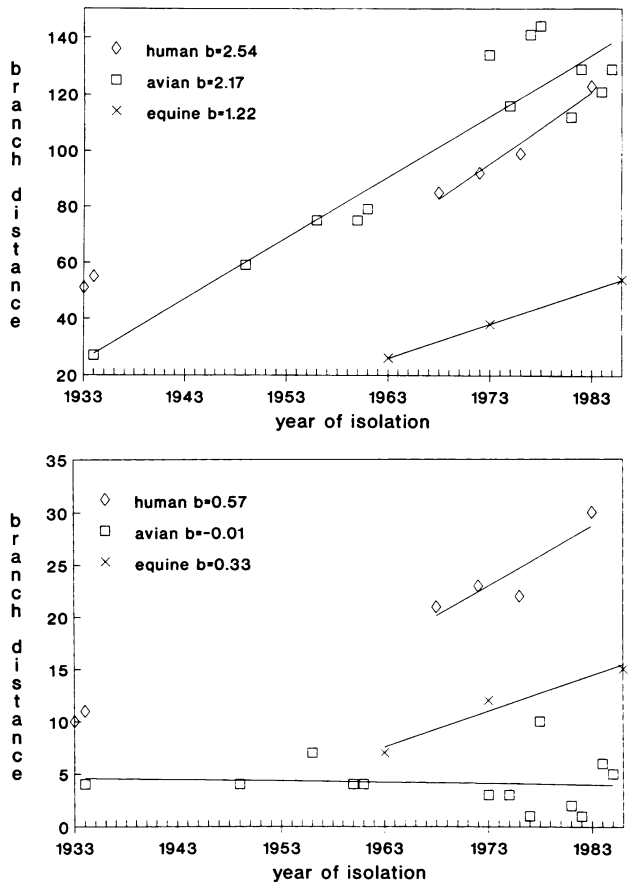

FIG. 5. Evolutionary rates for NP genes and proteins. (Top) NP nucleotide evolutionary rate. (Bottom) NP amino acid evolutionary rate. Evolutionary rate is estimated by regression of year of isolation against branch distance from the common ancestor node of the nucleotide and amino acid phylogenetic trees (Fig. 1). Regression statistic *b* provides rate estimates. The common ancestor node for the human lineage corresponds to the node dated 1933 in the nucleotide tree (Fig. 1, left). The common ancestor nodes for recent equine NPs and for Old World avian NPs are dated 1942 and 1925, respectively. Analogous common ancestor nodes of the amino acid tree are undated. The 1933 and 1934 human isolates (WS33, PR8-34) are not included in regression of human data.

American avian and H13 gull isolates were taken, we did not calculate evolutionary rates for these lineages. For the human lineage, accuracy could be improved by obtaining sequences for the 1934 to 1968 gap; we believe this would increase the total branch length of the lineage and the estimate of the evolutionary rate.

The estimation of evolutionary rates for amino acid sequences was similar (Fig. 5). For the human, recent equine, and avian lineages the rates were 0.57, 0.33, and −0.01 amino acid changes per year, respectively. Even though we obtained similar estimates for evolutionary rates at the nucleotide level for human and avian lineages, there appears to be no overall evolution at the amino acid level in the Old World avian lineage for more than 50 years. The rate of evolution at the protein level for the recent equine lineage is about half that for human NPs.

Estimation of dates for ancestral NPs. Estimates of evolutionary rates were used to calculate dates for hypothetical-ancestor nodes of the phylogenetic tree (Fig. 1, left). This was accomplished by dividing the branch-internodal distance by the evolutionary rate, yielding a distance in years. If the tree branching and distances are correct, and if evolutionary rates are accurate and constant within a lineage, then estimates of dates for ancestor nodes calculated from derivative branches should converge. Using this approach, we calculated dates for ancestor nodes for the recent equine, human, and Old World avian lineages (Fig. 1, left). The estimates of dates for ancestor nodes were unusually concordant in these lineages. The recent equine and human lineages predicted a date of 1860 to 1870 for the common ancestor of recent equine and human plus swine lineages. The combination of recent equine, human, and Old World avian lineages converges on the mid-19th century period of 1840 to 1860 for the origin of a common ancestor. The EQPR56 NP appears to have branched from an earlier ancestor. Using the relatively high Old World avian evolutionary rate, we conservatively estimated a pre-1800 date for the root node from which EQPR56 and all other NPs descend. From this analysis, we conclude that the present NP lineages are more than a century in age.

The only branch of the mammalian lineages that does not conform with the dates shown in Fig. 1 is that containing SWHK82 and SWTN77. Either the evolutionary rate is much reduced in this branch or the branch is misplaced; more classic H1N1 swine sequences between 1931 and 1977 may resolve this problem. The estimated common ancestor dates for these mammalian lineages suggest that H1N1 swine NPs and recent equine NPs may have been present in the late 19th century. The mammalian isolates MINKSW84, SWGER81, and SWNED85 appear to be evolving more slowly than other NPs in the Old World avian lineage; their branch distance points are below the regression line of Fig. 5. The date of the common ancestor node for the swine and avian isolates suggests that avian H1N1 NPs were present in swine at least 10 years before the date of virus isolation. At the nucleotide level, the SWHK76 isolate in the human lineage (Fig. 1) appears to be evolving at the same rate as other human isolates but at the amino acid level appears to be evolving more slowly (Fig. 5). These observations suggest that human and avian NPs evolve differently in a swine host.

Within the Old World avian lineage, the branch containing BDGRHO77 appears to be evolving faster than other contemporary NPs (e.g., DKHK75 and MLDAST82); their branch distance points are well above the regression line of Fig. 5. By using the evolutionary rate estimate for the Old World avian lineage, the projected dates for the common ancestors are 1912 for the Australia-New Zealand NPs (GTAUS79, DKNZ76), 1899 for the North American avian NPs, and 1882 for the H13 gull lineage (Fig. 1, left). The dates of these hypothetical ancestors do not coincide with the tree position or dates of isolation for the NPs in these three descendant branches. The discontinuity of dates in these other lineages relative to the Old World avian lineage may be the result of slower evolutionary rates or missing data, i.e., a lack of older isolates to accurately root these branches. For the North American and Australia-New Zealand NP lineages the discontinuity is correlated with geographic separation, but for the H13 gull lineage the discontinuity is correlated with a different separation, the limited host range of H13 viruses. Despite these discrepancies, the analysis shows that the evolutionary relationship of these other avian lineages to the Old World lineage is quite old; i.e., common ancestors must be dated well before the earliest Old World avian isolate (FPV34).

DISCUSSION

Host specificity and NP gene evolution. Phylogenetic trees represent hypotheses about evolutionary relationships among taxa. The results of our phylogenetic analysis identify the five host-specific NP RNA hybridization groups of Bean (1) as distinct lineages of NP genes. Furthermore, our analysis of amino acid sequences shows that these lineages are distinct at the protein level as well. Earlier studies have implicated the NP protein in determining the host range for influenza viruses (e.g., Scholtissek et al. [24]). This suggests that the NP protein may be involved in the evolution and maintenance of distinct gene pools for influenza A viruses. If this is true, then NP genes should fall into host-specific lineages, as we have demonstrated. Second, there must be adaptive functional differences at the protein level among host-specific NPs. This is suggested by the distinct amino acid differences we found among the lineages. Understanding of these differences must await detailed descriptions of the structure and function of the NP protein.

H13 NPs represent an unusual example of host specificity. Apparently influenza A viruses with H13 hemagglutinins are always coupled with a specific NP, and with the exception of one whale isolate, H13 NPs have not been found outside of gulls (10, 14). Viruses of other serotypes (H1, H2, H4, H11) bearing classic avian NPs have been isolated from gulls (1, 14). The specificity of the H13 NP suggests that it is highly adapted to gulls, but the basis for this adaptation is unknown.

Adaptation and rates of NP evolution. If an NP protein is well adapted to a host we might expect little, if any, evolution at the protein level to occur, but if an NP gene is introduced into a new host we might expect the NP protein to evolve rapidly. Mutations at the nucleotide level should fuel the evolutionary changes. In our analysis, human and Old World avian lineages have similar estimated rates of evolution at the nucleotide level, but at the amino acid level the human lineage is evolving faster than all other lineages and the Old World avian lineage appears not to be evolving at all. The evolutionary stasis of the Old World avian NP protein suggests that it is being maintained at an adaptive peak; the protein has not changed significantly for more than 50 years despite more than 300 nucleotide changes in the lineage. Also, there are fewer than 10 amino acid changes in many avian NP proteins relative to the root of the NP lineage, which is estimated to be more than a century in age. In contrast, NP proteins in other lineages continue to change, especially in the human lineage. This pattern is congruent with the hypothesis of an avian ancestor for all NPs and the hypothesis that the protein has not yet fully adapted to these new hosts. The basis for differences in evolution between avian and human NPs may be related to changes in host and tissue tropism (from gut endothelium in birds to respiratory epithelium in humans).

Reassortment and host fidelity of NP genes. If NPs are adapted to specific hosts, they should show host fidelity. This appears to be the case for most recent equine, H13 gull, and human NPs. Swine, however, readily accept viruses containing NP genes typical of human, avian, and classic swine lineages. The appearance of swine isolates in avian and mammalian lineages lends support to the mixing vessel hypothesis of Scholtissek et al. (24). The susceptibility of swine to avian H1N1 viruses and the ability of humans and swine to pass virus back and forth may have relevance to the 1918 influenza A virus pandemic; perhaps swine served as an intermediate step for the introduction of a new avianlike virus into the human population (as proposed by Scholtissek et al. [24]). Our analysis indicates that classic swine genes and genes introduced into the swine gene pool may be evolving slowly; this suggests that swine can serve as a stable reservoir for a variety of swine, human, and avian genes. Thus, swine appear to efficiently maintain a variety of genes which are available for production of new reassortants.

Common origin for classic swine and human NPs. The close relationship shown in our phylogenetic analysis between old human isolates (PR8-34 and WS33) and contemporary swine isolates (SWIA30 and SW31) compared with more recent human isolates strongly suggests a common origin for human and swine NPs. The estimated date for the ancestor node for these two groups is 1914, which is earlier than the 1918 pandemic of H1N1 swinelike flu (Fig. 1). Assuming a common ancestor for human and swine influenza viruses, Nakajima et al. (21) found similar results with the M and NS1 genes. Surprisingly, the current swine viruses containing H1N1-type NPs (SWTN77, SWHK82) do not appear to be direct descendants of SWIA30 or SW31 but do share a pre-1900 common ancestor. Perhaps the SWIA30 lineage was derived from the 1918 pandemic virus but is now extinct. This question can be resolved by sequencing classic swine NPs isolated between 1931 and 1977.

Geographic separation and NP gene evolution. The evolution of NP gene lineages appears to have been influenced by geographic separation of wild host populations. For example, the classic avian lineage is divided into North American, Australia-New Zealand, and Old World lineages (Fig. 1, left). The H13 lineage can be divided into Eurasian (GULAST84) and North American sublineages. For each of these geographic discontinuities, there appears to be a mismatch in the estimated date for the common ancestor and the dates of isolation for NPs in the descendant side lineages. For example, the Australia-New Zealand lineage containing the DKNZ76 NP appears to be older than FPV34, and NPs in this lineage appear to be evolving at less than half the rate of Old World avian NPs (Fig. 1, left). Donis et al. (8) have shown similar patterns of lineage divergence and reduced evolutionary rates correlated with geographic separation among H4 hemagglutinin genes.

Estimated ages of NP lineages. The time frame proposed for the evolution of the current diversity of NP lineages exceeds a century. This estimate is reasonable given the more than 50-year span over which viruses have been isolated from avian, swine, and human lineages. Our estimates for the age of NP lineages are likely to be conservative; the saturation of nucleotide sites with silent mutations and consequential undetected changes in old branches may result in smaller age estimates for older common ancestors. Thus we propose that the nonavian NP nucleotide lineages were well differentiated from the avian lineage by the end of the 19th century and that the EQPR56 lineage as well differentiated in the early 19th century.

Avian origin for influenza A virus NPs. Our phylogenetic analysis makes no assumptions about relationships among influenza A virus NPs. Independent evidence supporting the hypothesis of an avian (waterfowl) origin for influenza A viruses includes the proliferation and maintenance of numerous hemagglutinin and neuraminidase subtypes among migratory waterfowl populations, asymptomatic infection indicating coadaptation of virus and host, and a propensity for dissemination of viruses to other hosts because of migratory habits and the production of prodigious amounts of virus (11). In support of this hypothesis, most of the new influenza

virus genes that have appeared in mammalian gene pools over the past 30 years have been shown ultimately to have an avian origin (e.g., Hinshaw and Webster [11], Kida et al. [16], and Kawaoka et al. [15]). The results of our evolutionary analysis are consistent with the hypothesis of an avian origin for all NPs. Our summary evidence for this hypothesis includes the following: (i) older nonavian NP genes are more avianlike than recent NPs, (ii) recent NP genes in the five lineages are more distant from NPs in other lineages than from the most primitive avian NPs, (iii) nonavian NPs share more primitive avian characters than derived characters, and (iv) Old World avian NP proteins appear to be in evolutionary stasis while nonavian NPs are evolving rapidly away from avianlike ancestral NP proteins. These results support the hypothesis of independent evolution of NPs from an avian ancestor.

ACKNOWLEDGMENTS

We thank Evelyn Stigger, Raphael Onwuzuruigbo, and Daniel Channell for technical assistance, Clayton Naeve and the St. Jude Children's Research Hospital Molecular Resource Center for preparation of oligonucleotides, Patricia Eddy and the St. Jude Children's Research Hospital Molecular Biology Computer Facility for computer support, and David Swofford for assistance with PAUP software.

This work was supported by Public Health Service research grant AI-08831 and Public Health Service contract AI-52586 from the National Institute of Allergy and Infectious Diseases, Cancer Center Support (CORE) grant CA-21765, National Research Service Award 5T32-CA09346 to the senior author, and American Lebanese Syrian Associated Charities.

LITERATURE CITED

1. **Bean, W. J.** 1984. Correlation of influenza A virus nucleoprotein genes with host species. Virology **133**:438–442.
2. **Bean, W. J., G. Sriram, and R. G. Webster.** 1980. Electrophoretic analysis of iodine-labeled influenza virus RNA segments. Anal. Biochem. **102**:228–232.
3. **Briedis, D. J., and M. Tobin.** 1984. Influenza B genome: complete nucleotide sequence of the influenza B/Lee/40 virus genome RNA segment 5 encoding the nucleoprotein and comparison with the B/Singapore/222/79 nucleoprotein. Virology **133**:448–455.
4. **Buckler-White, A. J., and R. R. Murphy.** 1986. Nucleotide sequence analysis of the nucleoprotein gene of an avian and a human virus strain identifies two classes of nucleoproteins. Virology **155**:345–355.
5. **Buonagurio, D. A., S. Nakada, J. D. Parvin, M. Krystal, P. Palese, and W. M. Fitch.** 1986. Evolution of human influenza A viruses over 50 years: rapid and uniform rate of change in the NS gene. Science **232**:980–982.
6. **Chen, E. Y., and P. H. Seeburg.** 1985. Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. DNA **4**:165–170.
7. **Daniels, R. S., J. J. Skehel, and D. C. Wiley.** 1985. Amino acid sequences of haemagglutinins of influenza viruses of the H3 subtype isolated from horses. J. Gen. Virol. **66**:457–464.
8. **Donis, R. O., W. J. Bean, Y. Kawaoka, and R. G. Webster.** 1989. Distinct lineages of influenza virus H4 hemagglutinin genes in different regions of the world. Virology **169**:408–417.
9. **Gammelin, M., J. Mandler, and C. Scholtissek.** 1989. Two subtypes of nucleoproteins (NP) of influenza A viruses. Virology **170**:71–80.
10. **Hinshaw, V. S., W. J. Bean, J. Geraci, P. Fiorelli, G. Early, and R. G. Webster.** 1986. Characterization of two influenza A viruses from a pilot whale. J. Virol. **58**:655–656.
11. **Hinshaw, V. S., and R. G. Webster.** 1982. The natural history of influenza A viruses, p. 79–104. *In* A. S. Beare (ed.), Basic and applied influenza research. CRC Press, Inc., Boca Raton, Fla.
12. **Huddleston, J. A., and G. G. Brownlee.** 1982. The sequence of the nucleoprotein gene of human influenza A virus, strain A/NT/60/68. Nucleic Acids Res. **10**:1029–1038.
13. **Jones, K. L., J. A. Huddleston, and G. G. Brownlee.** 1983. The sequence of RNA segment 1 of influenza virus A/NT/60/68 and its comparison with the corresponding segment of strains A/PR/8/34 and A/WSN/33. Nucleic Acids Res. **11**:1555–1566.
14. **Kawaoka, Y., T. M. Chambers, W. L. Sladen, and R. G. Webster.** 1988. Is the gene pool of influenza viruses in shorebirds and gulls different from that in wild ducks? Virology **163**:247–250.
15. **Kawaoka, Y., S. Krauss, and R. G. Webster.** 1989. Avian to human transmission of the PB1 gene of influenza A viruses in the 1957 and 1968 pandemics. J. Virol. **63**:4603–4608.
16. **Kida, H., K. F. Shortridge, and R. G. Webster.** 1988. Origin of the hemagglutinin gene of H3N2 influenza viruses from pigs in China. Virology **162**:160–166.
17. **Krystal, M., R. M. Elliot, E. W. Benz, Jr., J. F. Young, and P. Palese.** 1982. Evolution of influenza A and B viruses: conservation of structural features in the hemagglutinin genes. Proc. Natl. Acad. Sci. USA **79**:4800–4804.
18. **Londo, D. R., A. R. Davis, and D. P. Nayak.** 1983. Complete nucleotide sequence of the nucleoprotein gene of influenza B virus. J. Virol. **47**:642–648.
19. **Mandler, J., and C. Scholtissek.** 1989. Localisation of the temperature-sensitive defect in the nucleoprotein of an influenza A/FPV/Rostock/34 virus. Virus Res. **12**:113–122.
20. **Nakada, S., R. S. Creager, M. Krystal, and P. Palese.** 1984. Complete nucleotide sequence of the influenza C/California/78 virus nucleoprotein gene. Virus Res. **1**:433–441.
21. **Nakajima, K., E. Nobusawa, and S. Nakajima.** 1984. Genetic relatedness between A/swine/Iowa/15/30 (H1N1) and human influenza viruses. Virology **139**:194–198.
22. **Reinhardt, U., and C. Scholtissek.** 1988. Comparison of the nucleoprotein genes of a chicken and a mink influenza A H10 virus. Arch. Virol. **103**:139–145.
23. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
24. **Scholtissek, C., H. Burger, O. Kistner, and K. F. Shortridge.** 1985. The nucleoprotein as a possible major factor in determining host specificity of influenza H3N2 viruses. Virology **147**:287–294.
25. **Scholtissek, C., I. Koennecke, and R. Rott.** 1978. Host range recombinants of fowl plague (influenza A) virus. Virology **147**:287–294.
26. **Snyder, M. H., A. J. Buckler-White, W. T. London, E. L. Tierney, and B. R. Murphy.** 1987. The avian influenza virus nucleoprotein gene and a specific constellation of avian and human virus polymerase genes each specify attenuation of avian-human influenza A/pintail/79 reassortant viruses for monkeys. J. Virol. **61**:2857–2863.
27. **Steuler, H., B. Schroder, H. Burger, and C. Scholtissek.** 1985. Sequence of the nucleoprotein gene of influenza A/parrot/Ulster/73. Virus Res. **3**:35–40.
28. **Tian, S. F., A. J. Buckler-White, W. J. London, L. J. Recle, R. M. Chanock, and B. R. Murphy.** 1985. Nucleoprotein and membrane protein genes are associated with restriction of replication of influenza A/mallard/NY/78 virus and its reassortants in squirrel monkey respiratory tract. Virus Res. **1**:625–630.
29. **Winter, G., and S. Fields.** 1981. The structure of the gene encoding the nucleoprotein of human influenza virus A/PR/8/34. Virology **114**:423–428.
30. **Winter, G., S. Fields, M. J. Gait, and G. G. Brownlee.** 1981. The use of synthetic oligodeoxynucleotide primers in cloning and sequencing segment 8 of influenza virus (A/PR/8/34). Nucleic Acids Res. **9**:237–245.